74251tra

Case 12/214 DI Fa/

BOEHRINGER INGELHEIM INTERNATIONAL GMBH

55216 Ingelheim am Rhein (BRD)

Tumour-associated antigen (B345)

The invention relates to the chemotherapy of tumour diseases.

Normal body cells are subject to a strictly regulated system which controls growth, cell division and the dying off of certain cells. Thus, an adult's body cells divide only when they are replacing dead cells or healing an injury. Cancer cells, on the other hand, carry on growing out of control, they accumulate and form a tumour. When a tumour reaches a critical size, cancer cells may be transported through the bloodstream or lymphatic system into other parts of the body and form colonies there (metastases). Not all tumours are cancerous, as benign tumours do not metastasise and are therefore not usually life-threatening as they can be surgically removed. More detailed information on this subject and the other aspects of tumour formation discussed hereinafter can be found in the following publications: Rauscher and Vogt, 1997; Kastan, 1997; Hesketh, 1995; Pusztai, Lewis and Yap, 1995; Wagener, 1996.

The transformation of a healthy cell into a cancer cell may be triggered by a whole range of factors such as environmental influences, radiation, viruses or chemical reagents. However, epigenetic modifications (methylations, acetylations and altered chromatin structure) and genetic modifications (point mutation, deletion, amplification, translocation) also play a crucial role in tumour development.

Mutations in coding regions of genes which are involved in regulating cell proliferation may contribute to the conversion of a normal cell into a tumour cell, as the transformed cell has growth advantages over its healthy neighbouring cell.

Cancer is therefore produced by an accumulation of inherited or acquired mutations in critical protooncogenes or tumour suppresser genes.

Cell proliferation is under the control of various gene systems, whereas products of oncogenes are involved in signal transmission from the cell surface to the cell nucleus, cyclin-dependent protein kinase and the inhibitors thereof accompany the cell through the cell cycle. Not infrequently, disorders in the synthesis of these proteins are found in tumour cells. The p53 protein plays a crucial role.

Proteins of the RB protein type regulate the availability of critical transcription factors.

The genes which are highly regulated in tumour tissues are usually the starting point for more detailed analyses and as proteins of all kinds of functions are highly expressed, the approach for therapeutic interventions may take many forms. The objective of cancer research is therefore to find new target molecules (so-called targets) for therapeutic interventions, which can then be used for a targeted therapy with few side effects.

The primary objective is therefore to detect molecular changes between normal tissue and tumour at the level

of gene expression ("transcription level"), which should identify new targets, on the one hand, and may be used to develop or find substances for inhibiting malfunctions, on the other hand.

A whole range of different methods of identifying and characterising new targets which form the starting point for developing new therapeutic agents are based on drawing up differential mRNA transcription profiles between tumours and normal tissues. These include differential hybridisation, establishing subtractions cDNA banks ("representational difference analysis"; Hubank and Schatz, 1994; Diatchenko et al., 1996) and the use of DNA chip technology or the SAGE method (Velculescu et al., 1995).

As well as immunotherapeutic approaches, targeted chemotherapy plays an essential role in the treatment of cancer. By chemotherapy is meant the administration of substances which have either a cytostatic or cytotoxic/cytolytic effect as a result of interfering with the metabolism, signal transduction and cell division processes of malignant cells. Chemotherapeutic agents can be divided into various categories on the basis of influencing specific targets in the tumour cell, the type of cellular interaction and interaction with a specific phase of the cell cycle.

The type of cancer treatment depends on the stage of the tumour, the critical point being whether metastases are already present and how far they have spread through the body. The administration of cell toxins for cancer treatment, as well as surgical intervention and

radiotherapy, is an integral part of present-day therapeutic concepts in oncology.

Chemotherapy has essentially two main aims: the primary one is to cure cancer; this means that the tumour disappears and does not return. If a cure is no longer possible for various reasons, attempts are made to restrict or control the growth and spread of the tumour.

In principle, substances used in chemotherapy are effective in all dividing cells. Tumour cells, however, are more sensitive to chemotherapeutic agents than healthy cells, as it is mainly strongly proliferating cells which are attacked.

Every tissue has its own growth characteristics, including cell division, stoppage of growth, differentiation and ageing, which are affected and regulated by internal and external factors.

Many of the cytotoxic chemotherapeutic agents currently used are effective only on proliferating cells (not in the G0 phase of cell division). Both normal and cancer cells are attacked. The destruction of normal cells may lead to severe side effects; e.g. destruction of the blood cell-producing tissues of the bone marrow (myelosuppression).

Chemotherapeutic agents are divided into various categories depending on how they affect specific substances within the tumour cell, the cellular processes with which the drugs interact and the phase of the cell cycle which they influence. This

information is necessary to oncologists in order to decide which preparations can be combined with one another in the therapy.

The highly regulated genes in tumour tissues are thus potential new target structures and, as proteins of all kinds of functions are highly expressed, this is a very versatile approach for therapeutic interventions.

The goal of cancer research is therefore to find new targets for therapeutic interventions which can then be used for targeted therapy with fewer side effects, compared with the therapeutic agents currently used.

In tumour tissues highly regulated genes are points of attack and hence potential target structures for chemotherapy.

The problem of the present invention was to provide a new protein, preferably expressed by tumour cells, which is a target molecule for intervention using chemotherapeutic methods.

This problem was solved by first producing a cDNA subtraction library, using RDA ("representational difference analysis") between a lung-adenocarcinoma cell line (A549) and normal lung tissue. In order to select the antigens overexpressed in the tumour the cDNA clones obtained were then sequenced and compared with sequences available in data banks. Among the annotated genes, there were 321 unknown ones, for which there were in most cases ESTs ("expressed sequence tags") -entries in the data bank. After further qualitative PCR analysis in cDNA libraries of critical

normal tissues and immunoprivileged tissues and more detailed data bank searches the number of candidate clones was restricted to 59, the ESTs of which do not come from critical normal tissues.

These clones were spotted onto Incyte DNA chips and hybridised with a whole range of tumour tissues and normal tissues as a reference. The mRNA expression profiles of EST fragments which are differentially expressed in cancer tissues and normal tissues and belong to an as yet unknown gene were verified using various methods.

The length of the transcripts was determined by Northern blot analysis and the expression pattern in different cell systems was exactly characterised by quantitative PCR. Only unknown genes or ESTs with tumour-specific expression profiles were followed up and subjected to "full length cloning". Potential ORFs ("open reading frames") are converted into the corresponding amino acid sequence and analysed for any possible prediction of function by in silico strategies.

Human B345-cDNA was cloned, and the sequence obtained is shown in SEQ-ID NO:1. Sequence analysis of the cloned human B345-cDNA showed that from position 215 to position 2461 (excluding stop codon) there is a continuous open reading frame which, at the nucleotide and protein level, shows no homology or identity in the known sequences of the data banks. The data obtained from the Northern Blot experiments lead one to conclude that the B345 transcript is about 6.5 kb long. The

cloned region of the B345-cDNA amounts to 5897 bp (excluding the polyA region), while the presence of a polyadenylation signal and the PolyA tail at the 3´-end of the sequence indicates that the cDNA is complete in this region. Because there is no continuous reading frame in the 5´ region of the cloned cDNA from position 1 to 214, it can be concluded that the ATG at position 215, which also corresponds 75% to a Kozak translation initiation site (ACCATGT) (Kozak, 1987), is the start codon of B345.

Additional information about the sequence of B345 located further upstream can be obtained by standard methods of molecular biology, e.g. by 5´-RACE ("rapid amplification of cDNA ends"). In this method, RNA, preferably mRNA, from cells or tissues in which B345 is transcribed (e.g. colon carcinoma tissue or cell line derived from a colon adenocarcinoma such as e.g. Colo205) is reverse transcribed and then ligated with an adapter of a known sequence. A PCR with an adapter primer (binds specifically to the adapter at the 5' end of the cDNA) and a B345-specific primer allows the amplification of corresponding B345 fragments. These PCR products can be cloned by standard methods and characterised, particularly by DNA sequencing.

An alternative method of characterising the 5´ end is to screen cDNA libraries by hybridising with DNA probes specific to B345. If the screening of cDNA libraries does not produce the desired result, on account of restrictions inherent in the method, e.g. inefficient reverse transcription caused by marked secondary structures of the RNA, genomic libraries may be

investigated, in the course of which clones containing the sequence information located upstream of the 5' end of the cDNA obtained, e.g. the promoter region of B345, may be isolated, for example, as in the screening of cDNA libraries, by hybridising with DNA probes specific to B345.

The isolated cDNA codes for the tumour-specific protein designated B345 with the amino acid sequence (B345) shown in SEQ-ID NO: 2. The sequence of B345 is defined by the start codon at position 215 of the isolated B345-cDNA.

Thus, in a first aspect, the invention relates to a tumour-specific protein designated B345, with the amino acid sequence shown in SEQ ID NO: 2.

The DNA molecules or fragments thereof according to the invention code for (poly)peptides designated B345 with the amino acid sequence shown in SEQ-ID NO: 2 or for protein fragments or peptides derived therefrom; this also includes DNA molecules which contain deviations from the sequence shown in SEQ-ID NO: 2 as a result of the degeneration of the genetic code.

In one embodiment of the invention, DNA molecules coding for the natural B345-polypeptide or for fragments thereof are used. Alternatively to the natural B345-cDNA or fragments thereof, modified derivatives may be used. These include sequences with modifications which code for a protein (fragment) or peptides with greater immunogenicity; the same considerations apply to the modifications at the DNA level as to the peptides described above. Another type

of modification is the lining up of numerous sequences, coding for immunologically relevant peptides, in the manner of a string of beads (Toes et al., 1997). The sequences may also be modified by the addition of auxiliary elements, e.g. functions which ensure more efficient release and processing of the immunogen (Wu et al., 1995). For example, by the addition of a locating sequence in the endoplasmatic reticulum ("ER targeting sequence") the processing and hence the presentation and, lastly, the immunogenicity of the antigen can be increased.

In another aspect the present invention relates to a recombinant DNA molecule which contains B345 DNA.

In another aspect the present invention relates to antibodies against B345 or fragments thereof. Polyclonal antibodies may be obtained in conventional manner by immunising animals, particularly rabbits, by injecting the antigen or fragments thereof, and then purifying the immunoglobulin.

Monoclonal anti-B345-antibodies may be obtained by standard procedures according to the principle described by Köhler and Milstein, 1975, by immunising animals, particularly mice, then immortalising antibody-producing cells of the immunised animals, e.g. by fusion with myeloma cells, and screening the supernatant of the resulting hybridomas by immunological standard assays for monoclonal anti-B345-antibodies. For therapeutic or diagnostic use in humans these animal antibodies may optionally be chimerised (Neuberger et al., 1984, Boulianne et al., 1984) or

humanised (Riechmann et al., 1988, Graziano et al., 1995) in the conventional manner.

Human monoclonal anti-B345-antibodies (fragments) may also be obtained from so-called "Phage Display Libraries" (Winter et al., 1994, Griffiths et al., 1994, Kruif et al., 1995, McGuiness et al., 1996) and using transgenic animals (Brüggemann et al., 1996, Jakobovits et al., 1995).

The anti-B345-antibodies according to the invention may be used in immunohistochemical analyses for diagnostic purposes.

According to another aspect the invention relates to the use of B345-specific antibodies for selectively delivering any desired substances to or into a tumour which expresses B345. Examples of such substances are cytotoxic agents or radioactive nuclides, the effect of which is to damage the area surrounding the tumour. Because of the relatively tumour-specific expression of B345 only mild side effects are to be expected. In another aspect substances for visualising tumours which express B345 may be used with the aid of B345-antibodies. This is useful for the diagnosis and for evaluating the course of therapy. Therapeutic and diagnostic uses for antibodies which may be used as anti-B345 antibodies are described in WO 95/33771.

The protein designated B345 according to the present invention and the protein fragments, peptides or peptide equivalents or peptidomimetics derived therefrom may be used in cancer therapy, e.g. to induce an immune response to tumour cells which express the

corresponding antigen determinants. They are preferably used for the treatment of B345-positive tumours, particularly in lung and colon carcinoma.

It is known that tumour-associated antigens may have tumour-specific mutations which contribute to an immunological differentiation between tumour and normal tissue (Mandruzzato et al., 1997; Hogan et al., 1998; Gaudi et al., 1999; Wölfel et al., 1994). In order to detect the presence of tumour-specific B345 mutations, appropriately using probes of the isolated cDNA according to the invention, the B345-cDNA from one or more different tumours is cloned and the sequences obtained are compared with normal tissue-B345-cDNA. Tests are carried out which are intended to show whether tumour-B345 peptides from a section of sequence which is mutated compared with normal tissue-B345 have greater immunogenicity than normal tissue-B345 peptides from the corresponding section. To confirm that some mutations are tumour-specific, antibodies against these regions may be generated and tumour cells may be investigated for the expression of possible mutations.

Thus, in another aspect, the present invention relates to B345-peptides, derived from regions of a tumour-expressed B345 which contain tumour-specific mutations.

It can be assumed, from the preferred expression of B345 in tumour cells, that this protein has an important function for the tumour, e.g. for its formation, infiltration and growth and thus constitutes a target for chemotherapeutic intervention.

With a view to its use as a target in targeted chemotherapy B345 is characterised in more detail so as to develop a suitable strategy for intervention with this function.

As the first step in the so-called "down-stream" functional analysis of B345, a bioinformatic analysis is conveniently carried out in a first step, to indicate the direction for the experimental validation of B345 as target.

For this analysis the bioinformatic concepts based on similarity and modular structure form an essential basis. Established bioinformatic tools for detecting similarities are BLAST (http://www.ncbi.nlm.nih.gov/BLAST, Altschul et al., 1997) or FASTA (Pearson & Lipman, 1988), the specialised data banks such as Pfam (http://www.sanger.ac.uk/Pfam, Bateman et al., 2000) and SMART (http://smart.embl-heidelberg.de, Schultz et al., 2000), which take account of domain structures. To refine the analysis applications such as Clustal (http://www2.ebi.ac.uk/clustalw, Higgins et al., 1996) HMMer (http://hmmer.wustl.edu), PSI-BLAST (Altschul et al., 1997) and the PROSITE data bank (http://www.expasy.ch/prosite, Hofmann et al., 1999) may be used. Statistical methods of analysis which are not based on homologies make it possible to predict other structurally and functionally relevant properties such as the secondary structure and the occurrence of transmembrane segments and helix-turn-helix motifs. There are methods of predicting the secondary structure of proteins; Jpred

(http://barton.ebi.ac.uk/servers/jpred.html, Cuff et al., 1998) is particularly worth mentioning. Predicting the secondary structure can underpin functional hypotheses, e.g. if the structure of the suspected homologue is known.

According to bioinformatic analysis B345 has a helical transmembrane domain, both the N-terminal and the C-terminal region being hydrophilic, leading to the conclusion that this protein is a transmembrane protein. The N-terminal, extracellular region has a few CUB domains which have a tendency to disulphide bridge formation and are therefore involved in dimerisation or protein-protein interactions (Bork et al., 1993). The C-terminal, intracellular end shows homology with a receptor kinase and a C-kinase substrate.

B345 is subsequently subjected to biochemical and biological analysis.

In a subsequent step the function of B345 for the progress of the tumour is clarified; e.g. by proliferation assays *in vitro* or in animal models which overexpress the B345 gene under investigation (constitutively or inducibly) and as a control either express it in deleted (inactive) form or regulate it down by antisense (cf. e.g. Grosveld and Kollias, 1992).

B345 may be used in screening assays to identify substances which modulate, particularly inhibit, the activity of this protein. In one embodiment an assay of this kind may consist, for example of introducing the B345 protein, or an active fragment thereof, into cells

which react to the activity of B345 with proliferation or of expressing the corresponding B345 cDNA in the cell, and determining the proliferation of the cells in the presence and absence of a test substance.

One example of test cells might be cells with a low division rate, e.g. primary cells containing no endogenous B345. To determine the suitability of the cells for a screening assay, the cells are transformed with B345-cDNA, cultured and tested using standard assays, e.g. the incorporation of thymidine, for their ability to proliferate. On the basis of a significant increase in their proliferation qualities after B345 expression they may be used as test cells, e.g. in High Throughput Screening Proliferation assays. Examples of proliferation assays in the High Throughput format, e.g. based on the MTS assay, are described in WO 98/00713.

Substances with a proliferation-inhibiting effect may be used to treat tumours with a significant B345 expression, particularly in carcinoma of the lung and colon.

Summary of the Figures:

Fig. 1A: expression profile of B345, B452 and B540 in individual lung carcinomas and lung tumour cell lines.

Fig. 1B: expression profile of B345 in normal large bowel tissue and tumour cell lines.

Fig. 1C: graphic representation of the alignment of B345, B452 and B540.

Fig. 2A: Northern Blot analysis of the tumour cell line A549 with a 490bp long B345 PCR product

Fig. 2B: Northern Blot analysis of various normal tissues with a 490bp long B345 PCR product

Fig. 2C: Northern Blot analysis of various cancer tissues with a 318bp long B345 PCR product

Fig. 3 mRNA expression analysis of B345 by real-time PCR of tumour and normal tissues.

Fig. 4: mRNA expression analysis of B345 by real-time PCR of laser-microscope-prepared large bowel tumours (LCM) and normal large bowel tissue and tumour cell lines.

Fig. 5: Graphic representation of the gene structure of B345.

Fig. 6: Hydrophilicity and transmembrane blot of the B345 protein

Fig. 7: Potential protein structure of B345

Description of the Tables

Tab. 1: Summary of the Northern Blot data for B345 in various normal tissues (1A) and cancer cell lines (1B)

Tab. 2A: Summary of the data for the quantitative PCR of B345 in various normal and cancer tissues

Tab. 2B: Summary of the data for the quantitative PCR of B345 in various normal tissues and microdissected colon adenocarcinoma tissues

Explanation of symbols

+++	extremely positive
++	strongly positive
+	positive
(+)	slightly positive
-	negative

Example 1

RDA ("Representational Difference Analysis") of the human adenocarcinoma cell line of the lung (A549) and normal lung tissue.

The human lung adenocarcinoma cell line A549 (CCL 185) obtained from ATCC was cultured in T150 cell culture flasks. The nutrient medium used was MEM with 10% heat-inactivated, foetal calf serum and 2 mM of L-glutamine. Every 3 to 4 days the cells were split for propagation by trypsinisation 1:5 to 1:10. After about 80% confluence had been achieved, 4 ml of a trypsin solution (contents per litre: 8g NaCl, 0.2g KCl, 1.13g anhydrous Na_2HPO_4, 0.2g KH_2PO_4, 100 ml of 2.5% trypsin solution, 1g EDTA-Na-salt; pH 7.2 - 7.4) were used per T150 cell culture flask to harvest the cells. The 4 ml were transferred into a 15 ml Falcon test tube, mixed with 8 ml of PBS, centrifuged at 1200 rpm in a Heraeus bench centrifuge (Megafuge 2.0R) for 5 min at 4°C, the cell pellet was mixed with 1 ml of lysing buffer (10 mM Tris-HCl pH8, 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% NP40), shaken vigorously and centrifuged off in a 2 ml Eppendorf vessel at 12,000 rpm and at 4°C for 5 min in a Sigma bench centrifuge (Sigma 202 MK). The supernatant was transferred into a new Eppendorf vessel and after the addition of 55 µl of 20% SDS solution extracted twice with double the volume of a $CHCl_3$/phenol (1:1 v/v) mixture and extracted once with a single volume of $CHCl_3$. The aqueous RNA-containing phase was mixed with 1/10 volume of 3M NaAc (pH5) and twice the volume of 96% EtOH and the RNA was precipitated overnight at -20°C. Starting from 1 mg of total RNA, the procedure for

isolating poly-A(+)RNA using the polyATtract Kit (Promega) was carried out according to the manufacturer's instructions. The A549 poly-A(+)RNA was stored in a concentration of 1 mg/ml in DEPC-treated H_2O in aliquots at -80^0C.

In order to carry out representational difference analysis (RDA; Hubank and Schatz, 1994; Diatchenko et al., 1996) the poly-A(+)RNA of the lung adenocarcinoma cell line A549 was used as the "tester" and that of normal lung tissue (1 mg/ml; Clontech, Palo Alto; #6524-1) was used as the "driver". The RDA was carried out using the PCR-select™ kit (Clontech, Palo Alto) in accordance with the manufacturer's instructions, except that a modified primer/adapter-2-oligonucleotide system was used: adapter-2-alt-1 (SEQ ID NO: 29) and nested-PCR-primer-2-alt (SEQ ID NO: 30) and adapter-2-alt-2 (SEQ ID NO: 31). The newly generated primer/adapter sequences make it possible to excise the particular cDNA fragments subsequently, thanks to the presence of three new restriction enzyme cutting sites (Kpn I, Sac I and Xho I) in the sequence of the nested-PCR-primer-2-alt after the subtracted cDNA fragments have been cloned into the pPCRII vector. It was necessary to design a primer/adapter sequence with a plurality of available restriction enzyme cutting sites because point mutations could often be observed, caused by the PCR amplification steps, in the primer sequences in particular.

After the synthesis of double-stranded cDNA using oligo-dT, the cDNA of "tester" and "driver" obtained was digested with *Rsa*I (RsaI is a restriction enzyme

which recognises 4 bases and on a statistical average yields cDNA fragments which are 256 bp long). Equal parts of "tester-cDNA" were ligated with either adapter 1 or 2 and then separately hybridised with an excess of "driver-cDNA" at 65°C. Then the two mixtures were combined and subjected to a second hybridisation with fresh denatured "driver cDNA". The concentrated "tester"-specific cDNAs were then exponentially amplified by PCR, with primers specific to the adapters 1 and 2. To achieve further concentration, one aliquot from this reaction was subjected to a second PCR with specific nested primers. The exponentially amplified cDNA fragments resulting from this reaction were ligated directly into the pCRII vector (Invitrogen; "TA-cloning vector")and then one third of the ligation mixture was transfected into competent *E. coli* (OneShot™, Invitrogen).

712 positive transformants (blue-white selection) were obtained and cultivated in 96-well blocks in LB-Amp Medium (1.3 ml per well) for 48 h at 37°C. 750 µl of the *E. coli* suspensions were used per well for the preparation of the plasmid-DNA (96-well QIAgen minipreparation method according to the manufacturer's instructions). The remaining bacterial cultures were stored as glycerol stock cultures at -80°C.

A cDNA subtraction library consisting of 712 individual clones was obtained, in the form of both *E. coli* glycerol stock cultures and also purified plasmids.

Example 2

DNA-sequencing and annotation of TAA candidates:

The isolated plasmid-DNA of all the 712 clones (cf. Example 1) was sequenced by the Sanger method on an ABI-377 Prism apparatus. The sequences obtained were annotated using BioScout-Software (LION, Heidelberg) and subjected to data bank comparisons (Genbank). Of 712 clones, 678 were able to be sequenced and annotated. The rest (34) either only had poly(A) sequences as an insert or corresponded to a religated vector or could not be sequenced. Of the 678 annotatable sequences, 357 proved to be genes with a known function. The remaining 321 represented clones coding for genes with an unknown function; 59 of them did not even have entries in the human EST data bank. Known genes were not treated further. For those unknown genes for which an EST entry was available, the expression profile was evaluated: all those ESTs with >95% identity (BLAST) which belonged to the experimentally determined sequence of the subtraction libraries were examined. During annotation the material was subdivided into a) critical normal tissue, b) foetal, "disposable" and immunoprivileged tissue and c) tumours and tumour cell lines. On the basis of this "virtual mRNA profile" ("virtual Northern blot") 200 clones for which no ESTs were found in group a) were selected for further experimental analyses (including the 59 clones for which there was no EST entry). To narrow down the candidate clones still

further, from the sequences determined from the 200 selected clones, pairs of oligonucleotide primers were designed and synthesised. First, 8 different cDNA libraries derived from human tissue (GibcoBRL "SUPERSCRIPT™"), which are directionally cloned in pCMV-SPORT, were tested by qualitative PCR for the presence of the particular candidates. The cDNA libraries used came from heart tissue (#10419-018), liver (#10422-012), Leukocytes (#10421-022), kidney (#10420-016), lung (#10424-018), testis (#10426-013), brain (#10418-010) and foetal brain (#10662-013). The PCR conditions were as follows: 20 µl of total volume per PCR mixture contained 1x TaqPol buffer(50mM KCl, 10 mM Tris-HCl pH 9, 0.1% Triton X-100), 1.5 mM $MgCl_2$, 0.2 mM dNTPs (Promega), 0.025 U/µl Taq-DNA-polymerase (Promega), 5 pM each of specific oligonucleotide primer for B345 (B345-D, SEQ ID NO: 32) and (B345-U, SEQ ID NO: 33) and 100 ng of the plasmid-DNA under investigation. Specific primers for GAPDH (SEQ ID NO: 34 and 35) were used as a control. To check the selective detection, the relevant B345 specific primer pairs, oligonucleotide primer (SEQ ID NO: 32) and (SEQ ID NO: 33), were tested in parallel for the isolated plasmid with the original B345 fragment (fragment of B345 cDNA originally isolated). The detectability of fragments of the expected length with a strong signal in one of the critical normal tissues (heart, liver, lung, kidney and leukocytes), but not in the cDNA libraries of immunoprivileged tissues (brain, foetal brain and testis) under these PCR conditions (1 cycle:

3´ 94°C; 35 cycles: 1´ 94°C - 1´ 55°C - 1´ 72°C; 1 cycle: 7´ 72°C) was defined as the distinguishing criterion. Using this qualitative PCR analysis the number of candidates could be reduced to 56; clone B345 was in this already preselected group of candidates.

Example 3

Expression analysis by cDNA chip hybridisation:

In order to design a cDNA chip a number of clones from categories having all kinds of functions ranging from apoptosis to cell cycle regulation were selected from the dBEST data bank by a nucleotide sequence search. In all, 1299 IMAGE clones were obtained (of which 1024 are known genes) and sequenced to check them. Microtitre plates with bacteria which contain approximately 800 bp long sequences from the 3` end of the gene in the vector were sent to Incyte Pharmaceuticals, Inc. (USA), where they were spotted onto 60 chips. In addition to these clones, 120 EST clones identified by RDA were also spotted onto the chips. The DNA chips thus produced were then hybridised with Cy3-labelled cDNA from normal tissue, tumour tissue and cell lines together with Cy5-labelled cDNA from a mixture of nine different normal tissues and the two signals were compared in order to standardise the expression values. The calculations were partly carried out in S-Plus or in Microsoft Excel. Evaluation of the chip experiments produced a very similar expression profile for B345,

B540 and B452 when hybridisation was carried out with lung cancer probes of cell lines and patient material (cf. Fig. 1A). A tumour-associated expression profile of this kind could also be found for B345 when colon adenocarcinoma was compared with normal colon tissue (cf. Fig. 1B).

The sequence alignment of B345, B540 and B452 clearly showed an overlap between the individual EST fragments. It could therefore be assumed that the three clones are ESTs of one and the same gene. The resulting DNA section covers a length of 843 bp (cf. Fig. 1C) and was used in further experiments to search public data banks. The search results shows no significant homology with known DNA or protein sequences, indicating that B345 is a hitherto unknown gene.

Example 4

Expression analysis of B345 using Northern Blots:

B345 is a gene of unknown function which is obviously highly regulated in tumour tissues (cf. Fig. 1A and 1B, Tab. 1a and Tab. 1B) according to DNA CHIP analyses.

In order to confer the transcription profile obtained, on the one hand, and determine the length of the expected mRNA for full size cloning, on the other hand, a Northern Blot analysis was carried out for B345 using human cell lines and the "Human Multiple Tissue Northern Blots" (Clontech and Invitrogen). The probes used were 490 bp and 318bp long PCR products of B345

(primer (SEQ-ID NO: 3 and SEQ-ID NO: 4 or SEQ-ID NO:5 and SEQ-ID NO:6)) labelled with [α-^{32}P]dCTP (NEN, Boston). The hybridisation took place at 68° for 2 h; visualisation by standard autoradiography (Hyperfilm, Amersham). Fig. 2A,2B and 2C and Tab. 1a and Tab. 1B show the results of this analysis: Fig. 2A the analysis of the cell line A549, Fig. 2B the analysis of 12 normal tissues (Peripheral Blood Lymphocytes (PBL), lung, placenta, small intestine, liver, kidney, spleen, thymus, colon, skeletal muscle, heart and brain) and Fig. 2C the analysis of 8 cancer cell lines (promyelocytic leukaemia HL60, HeLa-S3, chronic myelogenic leukaemia K-562, lymphoblastic leukaemia MOLT-4, Burkitt`s lymphoma (Raji), colon adenocarcinoma SW480, lung adenocarcinoma A549 and melanoma G361). The B345 transcript is 6.5 kb long.

Example 5

Analysis of the expression profile of B345 at the RNA level using quantitative RT-PCR (real time PCR or TaqMan analysis).

In order to quantify more precisely the expression of mRNA in the various normal and tumour tissues, "real time PCR" was used, which makes it possible to calculate the RNA concentration compared with an external standard.

The RNA was isolated from frozen tissue with Trizol according to the instructions provided by the manufacturer, Gibco. To eliminate any contaminating DNA

the prepared RNA was digested with DNAase I as follows: 3 µg of total RNA were incubated for 15 minutes at 37°C with 20 µl of 5× AMV buffer (Promega), 1 µl of RNasin (Promega) and 2 µl of DNase I (Boehringer Mannheim) in a total volume of 80 µl. 120 µl of phenol:chloroform: isoamyl alcohol (25:24:1) were added, mixed in a vortexer and briefly centrifuged. The aqueous phase was removed, 120 µl of chloroform:isoamyl alcohol (24:1) were added and the mixture was centrifuged as before. The purified RNA was precipitated with ethanol and dissolved in water.

Then the total RNA was transcribed into cDNA with reverse transcriptase (Superscript, Gibco, BRL): 1 µl of oligo dT primer (Promega) was added to 3 µg of total RNA and made up to a final volume of 10 µl with water. After 5 minutes' incubation at 70°C the solution was cooled for 5 minutes at room temperature. 5 µl of RT reaction buffer (5×, Gibco, BRL), 2.5 µl of dNTPs (10 mM of each, Boehringer Mannheim), 1 µl of RNasin (10U/µl, Promega), 1.5 µl of Superscript (10 U/µl, Gibco, BRL) and 5 µl of water were added and the mixture was incubated for 1 hour at 42°C and the reaction was stopped by incubating for 3 minutes at 95°C.

In order to prepare a cDNA pool of a specific type of tissue or tumour, 3 to 10 different individual preparations from different patients were mixed together in equal proportions.

The quantitative measurement of the "household genes" ß-actin, GAPDH and tubulin in cDNA pools was carried out as follows:

A) ß-Actin-TaqMan PCR (Perkin Elmer)

For details of the principle of the TaqMan method cf. manufacturer's information (Perkin Elmer). A TaqMan PCR run contained samples of ß-actin control sequence with 10^2, 10^3, 10^4, 10^5 and 10^6 copies/µl (Perkin Elmer) to determine the standard curve, a negative control without DNA and the cDNA pools which are to be quantified. All the samples were analysed in triplicate. For a 25 µl reaction mixture, 1 µl of cDNA, 2.5 µl of 10× buffer A (Perkin Elmer), 4 µl of $MgCl_2$ (25 mM, (Perkin Elmer)), 0.5 µl of each nucleotide (10 mM of dATP, dCTP, dGTP; 20 mM of dUTP), 0.125 µl of TaqMan probe (20 µM; TaqMan probe for ß-actin (SEQ-ID NO: 18 fluorescence-labelled with 6-carboxyfluorescein at the 5' end and with 6-carboxytetramethylrhodamine at the 3´ end), 1 µl of each ß-actin-specific primer (20 µM each of Forward primer SEQ-ID NO:19 and Reverse primer SEQ-ID NO:20), 0.25 µl of AmpErase uracil N-glycosylase "UNG" (1 U/µl, Perkin Elmer), and 0.125 µl of AmpliTaq Gold (5 U/µl, Perkin Elmer) were mixed together, transferred into MicroAmp Optical Tubes (Perkin Elmer) and sealed with MicroAmp Optical Caps. The PCR was carried out as follows: one cycle of 2 minutes at 50°C for the UNG reaction, one cycle of 10 minutes at 95°C to activate the AmpliTaq, 40 cycles each of 15 seconds at 95° and 1 minute at 60°C. Then the probes were kept at 25°C. The data was evaluated using the "Sequence Detection System 1.5b1" programme

(PE Applied Biosystems), basically by comparing the fluorescence signals of the cDNA probes being quantified with the signals of the control plasmid dilutions of known concentration.

B) GAPDH-TaqMan PCR

For quantifying GAPDH, which was used like ß-actin or tubulin to standardise the RNAs used, the following primers or probes were used. The TaqMan probe used for GAPDH was a probe (SEQ-ID NO: 21) labelled at the 5' end with tetrachlorofluorescein and at the 3' end with carboxymethylrhodamine (Forward GAPDH primer: SEQ-ID NO: 22 and Reverse primer: SEQ-ID NO: 23). The reactions were carried out as described above.

C) Tubulin-SybrGreen PCR (Perkin Elmer)

For the principle of SybrGreen PCR cf. the manufacturer's information (Perkin Elmer). A SybrGreen PCR run contained samples of tubulin control plasmid with 10^2, 10^3, 10^4, 10^5 and 10^6 copies/μl (Perkin Elmer) to determine the standard curve, a negative control without DNA and the cDNA pools which are to be quantified. All the samples were analysed in triplicate. For a 25 µl reaction mixture, 1 µl of cDNA, 2.5 µl of 10× SybrGreen buffer (Perkin Elmer), 3,5 μl $MgCl_2$ (25 mM, Perkin Elmer)), 0.5 μl of each primer (20 μM each, Perkin Elmer, Tubulin Forward (SEQ-ID NO:24); Tubulin reverse (SEQ-ID NO:25), 0.25 μl of AmpErase uracil N-glycosylase "UNG" (1 U/μl, Perkin Elmer), and 0.25 μl of AmpliTaq Gold (5 U/μl, Perkin

Elmer) were mixed together, transferred into MicroAmp Optical Tubes (Perkin Elmer) and sealed with MicroAmp Optical Caps. The PCR was carried out as follows: one cycle of 2 minutes at 50°C for the UNG reaction, one cycle of 10 minutes at 95°C to activate the AmpliTaq, 40 cycles each of 15 seconds at 95° and 1 minute at 60°C. Then the probes were kept at 25°C. The data was evaluated using the "Sequence Detection System 1.5b1" programme (PE Applied Biosystems), basically by comparing the fluorescence signals of the cDNA probes being quantified with the signals of the control plasmid dilutions of known concentration.

D) B345-TaqMan PCR

The quantitative TaqMan-PCR analysis of B345 was carried out as described for the "household genes". However, B345 specific primers (SEQ-ID NO:26 and SEQ-ID NO:27) (200 ng/μl) and a B345 specific probe (SEQ-ID NO:28, 20 μM) labelled at the 5' end with tetrachlorofluorescein and at the 3' end with carboxymethylrhodamine were used. The PCR product of B345 with the primers SEQ-ID NO: 26 and SEQ-ID NO: 27 with a known copy number was used as the standard.

Fig. 3 shows that B345 is expressed more highly in large bowel cancer tissue than in normal tissue (cf. Tab. 2a). However, both the normal tissue and the tumour tissue constitute a very heterogeneous mixture of different cell types. Furthermore, the proportion of tumour cells in the tumour tissue varies considerably from about 30 to 80%. In order to minimise this

biological heterogeneity, the epithelial cells of the large intestine, which are the cells of origin for the adenocarcinoma, and cancer cells or cancer regions were specifically prepared by laser microdissection. Tissue sections 10μm thick were prepared using a cryomicrotome made by Leica, Jung CM1800 and placed on a polyethylene-coated slide (Böhm et al., 1997). The sections, dried at ambient temperature for about 30 minutes, were incubated with Mayers haematoxylin (SIGMA DIAGNOSTICS) and then washed under running water for five minutes to remove any non-specifically bound dye. After 5 minutes' drying at 37°C, the laser microdissection was carried out. This was done using the PALM laser microscope (PALM GmbH, Bernried, Germany) and about 2000 to 5000 cells were prepared. The cDNA obtained by Reverse Transcription was in turn analysed by Real Time PCR. The results show that the B345 expression in large bowel carcinoma cell lines and in patient material is comparatively many times higher than that of the normal tissue of the large bowel. For standardisation, the expression level of GAPDH was determined (cf. Fig. 4 and Tab. 2B).

Example 6

Cloning of the total cDNA of B345

There are a number of data banks with sequences of gene fragments (ESTs, expressed sequence tags) which can be used for the "in silico" cloning of genes and which,

when gone through with B345, yielded an overlapping EST contig of about 1500 bp. The polyA region at one of the ends indicated the orientation of the DNA section in relation to 5' - 3' orientation, which is essential when designing new primers for the amplification of B345-specific cDNA fragments.

First, the potential 3' end described by the data bank analysis was verified by experimental approaches. RNA from the lung carcinoma cell line Calu 6 (AACC No. HTB56) was reverse transcribed using the primer (SEQ-ID NO:7) and the resulting single-strand cDNA was amplified by PCR with the gene-specific primer SEQ-ID NO:3 and the adapter primer SEQ-ID NO: 8.

For a 25 µl PCR mixture, 1 µl of the cDNA pool was mixed in water with 2.5 µl of 10×Taq buffer (Promega), 1.5 µl of $MgCl_2$ (25 mM, Promega), 0.5 µl of dNTPs (10 mM each, Boehringer Mannheim), 1 µl of primer mixture (20 µM each), 0.15 µl of Taq polymerase (Promega). The PCR was carried out as follows: 1× 94°C for 3 minutes; 30× 94°C for 30 seconds, 55°C for 30 seconds, 72°C for 1 minute; kept at 4°C. The PCR was analysed on a 1.2% agarose gel.

The two primers were then used to sequence the purified PCR product. The sequences found showed high homology with the DNA fragment cloned "in silico" (including the poly-tract).

Since the cloning of 5' terminal sequences is usually a very laborious process, various methods were used below to solve the problem.

Once again, Calu 6 was used as the starting cell line. After reverse transcription of the RNA with the primer SEQ-ID NO: 7 and synthesis of the second strand, a linker consisting of the two oligos SEQ-ID NO: 9 and SEQ-ID NO: 10 was ligated onto the double stranded cDNA (Abe et al., 1992). The resulting LoneLinker cDNA library was then amplified with the gene-specific primer SEQ-ID NO: 4 in linear manner over 35 cycles. One aliquot of the B345-enriched cDNA could then be further amplified with the primers SEQ-ID NO: 11 and LLEcoRIA SEQ-ID NO: 9. After gel electrophoresis of one aliquot and Southern analysis with the gene-specific oligo SEQ-ID NO: 12, it was possible to locate a 5 kb band. This fragment was then sequenced step by step and aligned with the EST contig.

In order to check the resulting sequence from the LLcDNA cloning, two fragments were amplified by PCR (SEQ ID NO: 13 and SEQ ID NO: 14 or SEQ ID NO: 4 and SEQ ID NO: 15) used to screen Lambda gt10 cDNA phage libraries. Positive plaques were isolated and amplified by PCR using gt10-specific primers (SEQ ID NO: 16 and SEQ ID NO: 17). Subsequent sequencing and alignment with the sequences led to the assumption that this is a differential splice product. The splice donor, acceptor and lariat sequence were found subsequently. Using PCR by a suitable combination of primers, searches were carried out in various cell lines for differential splice products; in all the cell lines screened, only one product was found and led to the gene structure shown in Fig. 5. The cDNA shown has an open reading frame (ORF) which codes for a potential protein 749

amino acids long. The translation initiation site at position 215 approximately corresponds to 75% of a Kozak consensus sequence. However, the translation initiation site still has to be determined exactly by primer extension or RNAse protection, in order to be certain that the 5' end of B345 is present. The amino acid sequence of B345 is shown in SEQ-ID NO: 2.

Example 7

Bioinformatic analysis: Possible function of B345

The resulting primary amino acid sequence of B345 is shown in SEQ ID NO: 2. Analysis of the hydrophilicity plot of the amino acid sequence using the method of Kyte and Doolittle (1982) with a window size of 7 shows that the B345 protein has a characteristic hydrophobic domain (AS 600-622), which can be assumed to be a helical transmembrane domain (cf. Fig. 6). Both the C-terminal and N-terminal ends of B345 are hydrophilic. This polarised structure indicates that in all probability this is an integral membrane protein. The transmembrane helix connects an extracellular section about 600 amino acids long and a short intracellular section (125 amino acids) (cf. Fig. 7).

The extracellular domain also shows clear signs of the existence of a CUB domain at positions 158-247 and less clear signs of 2 other CUB domains in the region of 360-600. The C-terminal intracellular region might possibly be a kinase substrate. The CUB domains occur in various proteins, generally in developmentally

regulated proteins. In addition, CUB domains can sometimes also be found at EGF-like domains. The protein also has 12 potential N-glycosylation sites which can be found in the predicted extracellular domain, which again concurs with the predicted orientation of the protein.

With a BLAST hit (E-value: 5.8×10^{-2}) for the region from 169 to 216 of B345 it was possible to identify a complement-activating component of the *RA-reactive factor* (RARF) from *mus musculus*. The alignment is located within the CUB domain 1 of B345.

The CUB domains 2 and 3 (section 359-471 and 481-594) exhibit marginal homology with the human and fugu procollagen C proteinase enhancer protein (PCOLCE). These regions occur in the part of PCOLCE which contains a CUB domain tandem repeat (E-values: 0.5 (human) and 2.7 (fugu)). CUB domains sometimes occur in repeats.

Presumably, the B345 protein form a ß-sheet secondary structure, as it is known that CUB domains fold into a ß sandwich.

The intracellular domain (section 623-749) has no significant homologies. However, the total C-terminus aligned with an EST of human ovarian cancer cells (82% identity over 124 amino acids). The region from 602-656 shows homology with a possible A. thaliana receptor kinase, whereas the region around positions 611-659 shows similarities to an A. thaliana potential serine/threonine kinase.

Tab.1A

tissue	expression
PBL	-
lung	++
placenta	+
small intestine	+
liver	-
kidney	++
spleen	-
thymus	-
colon	+
skeletal muscle	-
heart	-
brain	-

Tab.1B

cell line	expression
promyelocytic leukaemia HL60	-
HELA cells S3	-
chronic myelogenic leukaemia K-562	+
lymphoblastic leukaemia MOLT-4	-
Burkitt`s lymphoma (Raji)	-
colon adenocarcinoma SW480	+++
lung adenocarcinoma A549	+
melanoma G361	-

Tab.2A

tissue	expression B345 / actin	expression B345 / tubulin
lung adenocarcinoma	+	+
lung adenocarcinoma	+	+
normal lung	- to (+)	(+)
colon adenocarcinoma	++	++
colon adenocarcinoma	+++	+++
normal colon	- to (+)	+
mammary IDC	+	+
breast	-	-
Hodgkin`s lymphoma	-	-
spleen	-	-
testis	-	-

Tab.2B

cell lines and tissue	expression B345 / GAPDH
colon adenocarcinoma SW480	+
normal colon (Clontech)	(+)
normal colon (Invitrogen)	(+)
lung adenocarcinoma A549	(+)
colon adenocarcinoma Colo 205	+++
PG 102142 tumour (Colon Ac.)	+++
PG 21900 tumour (Colon Ac.)	++
PG 7066 tumour (Colon Ac.)	+++
PG 32389 tumour (Colon Ac.)	++

Bibliography

Abe, K., Rapid isolation of desired sequences from lone linker PCR amplified cDNA mixture: application to identification and recovery of expressed sequences in cloned genomic DNA. Mamm. Genome 2,252-259

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402 (1997).

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. and Sonnhammer, E. L. The Pfam Protein Families Database. Nucleic Acids Res. 28, 263-266 (2000).

Bork, P. et al. (1993). The CUB domain. A widespread module in developmentally regulated proteins. J.Mol. Biol. 231: 539-545

Boulianne, G. L., et al., (1984), Nature 312: 643-646

Böhm et al., A.J. of Pathology 151,1:63-67, 1997

Brüggemann, M. and Neuberger, M.S., (1996), Immunol. Today 17: 391-397

Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. Jpred: a consensus secondary structure prediction server. Bioinformatics 14, 892-893 (1998).

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., and Siebert, P.D. (1996), Proc. Natl. Acad. Sci. U. S. A. 93, 6025-6030.

Gaudi C, Kremer F, Angevin E, Scott V, Triebel F (1999), J Immunol 162:1730-1738

Graziano, R.F., et al., (1995), J. Immunol. 155: 4996-5002

Griffiths, A.D., et al., (1994), EMBO J. 13: 3245-3260

Grosveld, F. and Kollias, G. Transgenic Animals, Academic Press (1992)

Hesketh, R., (1995), The oncogene, Academic Press

Higgins, D. G., Thompson, J. D. and Gibson, T. J. Using CLUSTAL for Multiple Sequence Alignments. Methods Enzymol. 266, 383-402 (1996).

Hogan KT, Eisinger DP, Cupp SBC, Lekstrom KJ, Deacon DD, Shabanowitz J, Hunt DF, Engelhard VH, Slingluff CL, Ross MM (1998), Cancer Res 58:5144-5150

Hofmann, K., Buchner, P., Falquet, L. and Bairoch, A. The PROSITE database, its status in 1999. Nucleic Acids Res. 27, 215-219 (1999).

Hubank, M. and Schatz, D.G., (1994), Nucleic. Acids. Res. 22, 5640-5648.

Jakobovits, A., (1995), Curr. Opin. Biotechnol. 6: 561-566

Kasten, M.B., (1997), Genetic Instability and Tumorigenesis, Springer Verlag

Köhler, G. and Milstein, C. (1975), Nature 265, 495-497

Kozak, M (1987), An analysis of 5`noncoding sequences from 99 vertebrates messenger RNA´s . Nuc.Ac.Res. Vol.15: 8125-8147

Kruif, J., et al., (1995), Proc. Natl. Acad. Sci. USA 92: 3938-3942

Kyte, J and Doolittle, RF (1982), J. Mol. Biol. 157: 105-132

Mandruzzato S, Brasseur F, Andry G, Boon T, van der Bruggen P (1997) , J Exp Med 186:785-793

McGuinnes, B.T., et al., (1996), Nature Biotechnol. 14, 1149

Neuberger, M.S., et al., (1984), Nature 312: 604-608

Pearson, W. R. and Lipman, D. J. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444-2448 (1988).

Pusztal et al., 1996, cell proliferation in cancer, Oxford medical publications

Rauscher, F.J. et al., (1997), Chromosomal translocations and oncogenic transcription factors, Springer Verlag

Riechmann, L., et al., (1988), Nature 332: 323-327

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. SMART: A Web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231-234 (2000).

Toes, R.E., Hoeben, R.C., Van der Voort, E., Ressing, M.E., Van-der-Eb, A.J. Melief, C.J.M., and Offringa, R. (1997), Proc. Natl. Acad. Sci. U.S.A. 94 (26): 14660-14665

Velculescu, VE, Zhang, L, Vogelstein, B, and Kinzler, KW (1995), Science 270: 484-487

Wagner, C., 1996, Molekulare Onkologie, Thieme

Winter, G., et al., (1994), Annu. Rev. Immunol. 12, 433-455

Woelfel, T, Schneider, J, Zum Buschenfelde, Meyer, KH, Rammensee, HG, Rotzschke, O, and Falk, K (1994), Int. J. Cancer 57: 413-418

Wu, T.C., Guarnieri, F.G., Staveley-O'Carroll, K.F., Viscidi, R.P., Levitsky, H.I., Hedrick, L., Cho, K.R., August, J.T., and Pardoll, D.M. (1995), Proc. Natl. Acad. Sci. U.S.A. 92(25): 11671-11675.